

# Regresión no Lineal

Cristian López Del Alamo  
[clopezd@utec.edu.pe](mailto:clopezd@utec.edu.pe)  
IPRODAM3D - Research group

2022

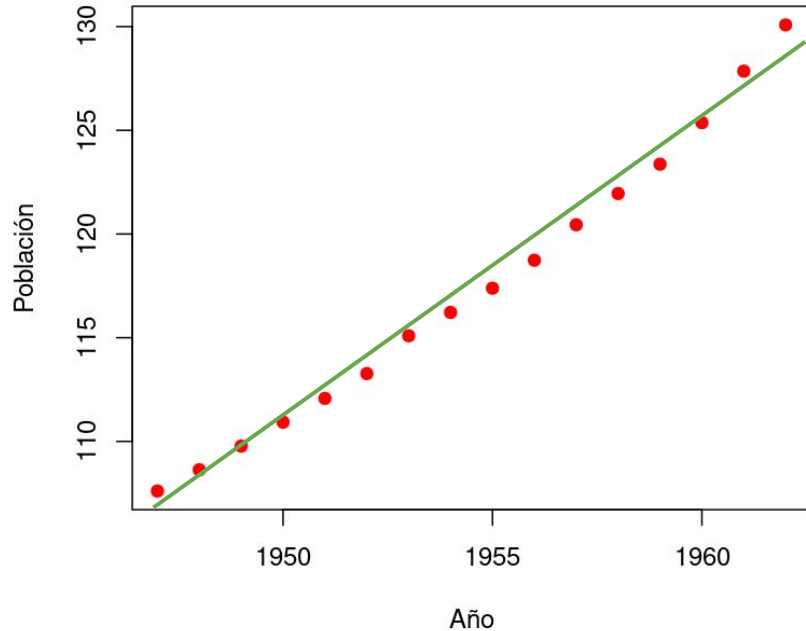
# 1

## Recordando: Regresión no Lineal

**Objetivo:** En esta clase es entender la similitud entre la regresión lineal y la regresión no lineal, las aplicaciones y el algoritmo de aprendizaje de máquinas que está relacionado.



# Regresión Lineal Simple



Hipótesis

$$h(x_i) = b + wx_i$$

Función de Pérdida

$$\mathcal{L} = \frac{\sum_{i=0}^n (y_i - h(x_i))^2}{2n}$$

Cálculo de Derivadas

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{\sum_{i=0}^n (y_i - h(x_i))(-x_i)}{n}$$

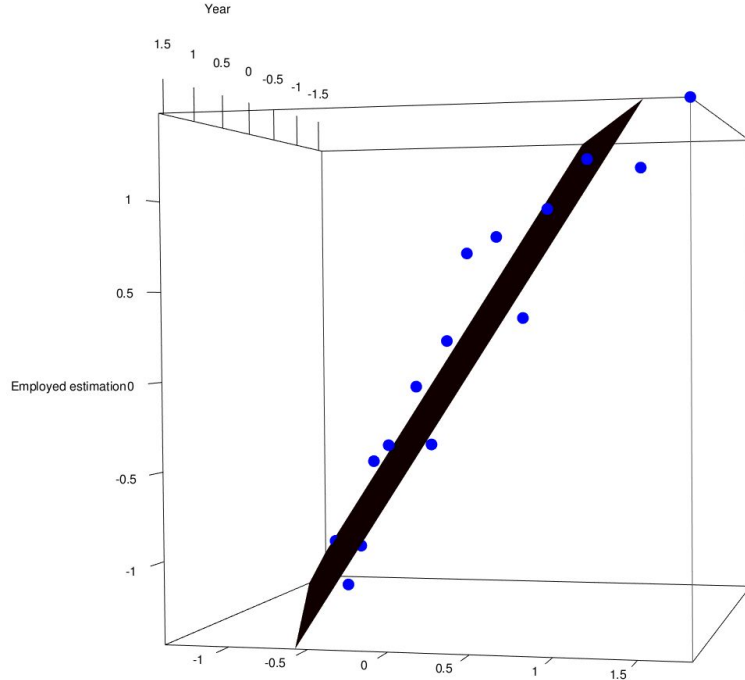
$$\frac{\partial \mathcal{L}}{\partial b} = \frac{\sum_{i=0}^n (y_i - h(x_i))(-1)}{n}$$

Actualización

$$w = w - \alpha \frac{\partial \mathcal{L}}{\partial w}$$

$$b = b - \alpha \frac{\partial \mathcal{L}}{\partial b}$$

# Regresión lineal múltiple: Caso 2 variables



Hipótesis

$$h(x_i) = b + x_{(i,1)}w_1 + x_{(i,2)}w_2$$

Función de Pérdida

$$\mathcal{L} = \frac{\sum_{i=0}^n (y_i - h(x_i))^2}{2n}$$

Cálculo de Derivadas

$$\frac{\partial \mathcal{L}}{\partial w_j} = \frac{\sum_{i=0}^n (y_i - h(x_i))(-x_{(i,j)})}{n}$$

$$\frac{\partial \mathcal{L}}{\partial b} = \frac{\sum_{i=0}^n (y_i - h(x_i))(-1)}{n}$$

Actualización

$$w_j = w_j - \alpha \frac{\partial \mathcal{L}}{\partial w_j}$$

$$b = b - \alpha \frac{\partial \mathcal{L}}{\partial b}$$

# Regresión lineal múltiple: Caso k variables

## Hipótesis

$$h(x_i) = b + x_{(i,1)}w_1 + x_{(i,2)}w_2 + \dots + x_{(i,k)}w_k$$

$$h(x_i) = 1w_0 + x_{(i,1)}w_1 + x_{(i,2)}w_2 + \dots + x_{(i,k)}w_k$$

$$h(x_i) = x_{(i,0)}w_0 + x_{(i,1)}w_1 + x_{(i,2)}w_2 + \dots + x_{(i,k)}w_k$$

$$h(x_i) = [x_{(i,0)} \ x_{(i,1)} \ x_{(i,2)} \ \dots \ x_{(i,k)}][w_0 \ w_1 \ w_2 \ \dots \ w_k]^T$$

$$h(x_i) = x_i * w^T$$

# Regresión lineal múltiple: Caso k variables

Hipótesis

$$\begin{array}{c}
 \text{Matrix } X \\
 \begin{array}{c} 0 \quad 1 \quad \dots \quad k \\
 \begin{array}{|c|}
 \hline
 \begin{array}{c}
 0 \quad 1 \quad \dots \quad x_{0k} \\
 1 \quad x_{11} \quad \dots \quad x_{1k} \\
 \vdots \quad \vdots \quad \dots \quad \vdots \\
 \boxed{i \quad 1 \quad x_{i1} \quad \dots \quad x_{ik}} \\
 \vdots \quad \vdots \quad \dots \quad \vdots \\
 n \quad 1 \quad x_{n1} \quad \dots \quad x_{nk}
 \end{array}
 \end{array}
 \end{array}
 \end{array}
 \times
 \begin{array}{c}
 \text{Vector } w \\
 \begin{array}{|c|}
 \hline
 \begin{array}{c}
 0 \quad w_0 \\
 1 \quad w_1 \\
 \vdots \quad \vdots \\
 k \quad w_k
 \end{array}
 \end{array}
 \end{array}
 =
 \begin{array}{c}
 \text{Vector } \bar{y} \\
 \begin{array}{|c|}
 \hline
 \begin{array}{c}
 0 \quad h(x_0) \\
 1 \quad h(x_1) \\
 \vdots \quad \vdots \\
 i \quad h(x_i) \\
 \vdots \quad \vdots \\
 n \quad h(x_n)
 \end{array}
 \end{array}
 \end{array}$$

$$h(X) = X * w^T$$

# Regresión lineal múltiple: Caso k variables

Función de Pérdida

$$\begin{array}{c} \text{Vector} \\ \textcolor{red}{y} \\ \begin{array}{c} 0 \\ 1 \\ \vdots \\ i \\ \vdots \\ n \end{array} \begin{array}{c} y_0 \\ y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{array} \end{array} - \begin{array}{c} \text{Vector} \\ \textcolor{red}{\bar{y}} \\ \begin{array}{c} 0 \\ 1 \\ \vdots \\ i \\ \vdots \\ n \end{array} \begin{array}{c} h(x_0) \\ h(x_1) \\ \vdots \\ h(x_i) \\ \vdots \\ h(x_n) \end{array} \end{array} = \begin{array}{c} \textcolor{red}{y} \\ \begin{array}{c} 0 \\ 1 \\ \vdots \\ i \\ \vdots \\ n \end{array} \begin{array}{c} y_0 - h(x_0) \\ y_1 - h(x_1) \\ \vdots \\ y_i - h(x_i) \\ \vdots \\ y_n - h(x_n) \end{array} \end{array} \longrightarrow \begin{array}{c} \textcolor{red}{y} \\ \begin{array}{c} 0 \\ 1 \\ \vdots \\ i \\ \vdots \\ n \end{array} \begin{array}{c} (y_0 - h(x_0))^2 \\ (y_1 - h(x_1))^2 \\ \vdots \\ (y_i - h(x_i))^2 \\ \vdots \\ (y_n - h(x_n))^2 \end{array} \end{array}$$

sum  $= (y_i - h(x_i))^2 / 2n$

$$\mathcal{L} = \|y - h(X)\|_2^2$$



# Regresión lineal múltiple: Caso k variables

Cálculo de derivadas

Handwritten matrix representation of the loss function derivative calculation:

Left matrix (Residuals):

$$\begin{matrix} & y - \hat{y} \\ 0 & y_0 - h(x_0) \\ 1 & y_1 - h(x_1) \\ \vdots & \vdots \\ i & y_i - h(x_i) \\ \vdots & \vdots \\ n & y_n - h(x_n) \end{matrix}$$

Right matrix (Features):

$$\begin{matrix} & 0 & 1 & \dots & j & \dots & K \\ 0 & 1 & x_{01} & \dots & -x_{0j} & \dots & x_{0K} \\ 1 & 1 & x_{11} & \dots & -x_{1j} & \dots & x_{1K} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ i & 1 & x_{i1} & \dots & -x_{ij} & \dots & x_{iK} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ n & 1 & x_{n1} & \dots & -x_{nj} & \dots & x_{nK} \end{matrix}$$

Note: The column labeled 'j' is highlighted with a red box, and the entries  $-x_{ij}$  are marked with green dots. The column header 'j' is also written in green above the box.

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{(Y - h(X))^T * (-1 * X)}{n}$$

# Regresión lineal múltiple: Caso k variables

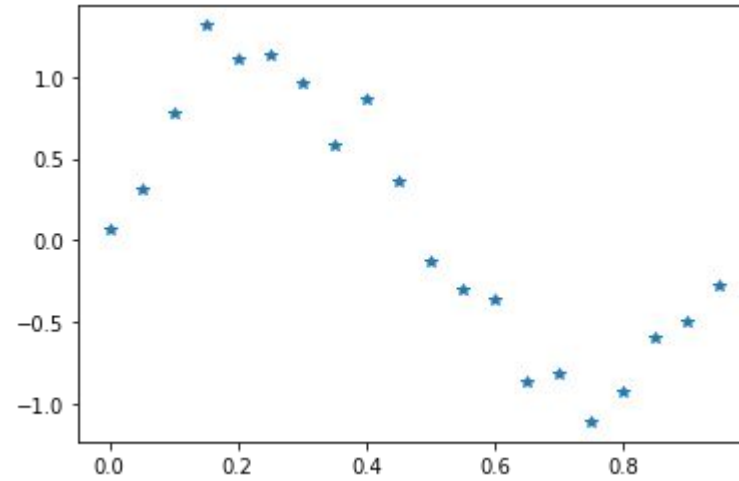
Actualizando w

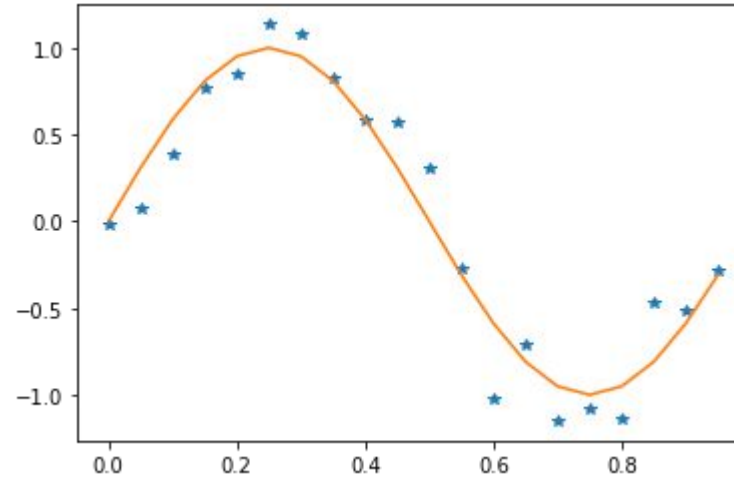
$$W = W - \alpha * \frac{\partial \mathcal{L}}{\partial w}$$

# 2

## Regresión no Lineal

**Objetivo:** Entender la matemática detrás de la regresión no lineal multivariable







## Polinomio

hipótesis:  $h(x_i) = b + x_i w_1 + x_i^2 w_2 + x_i^3 w_2 + \dots + x_i^p w_2$

hipótesis:  $h(x_i) = b + \sum_{j=1}^p x_i^j w_j$

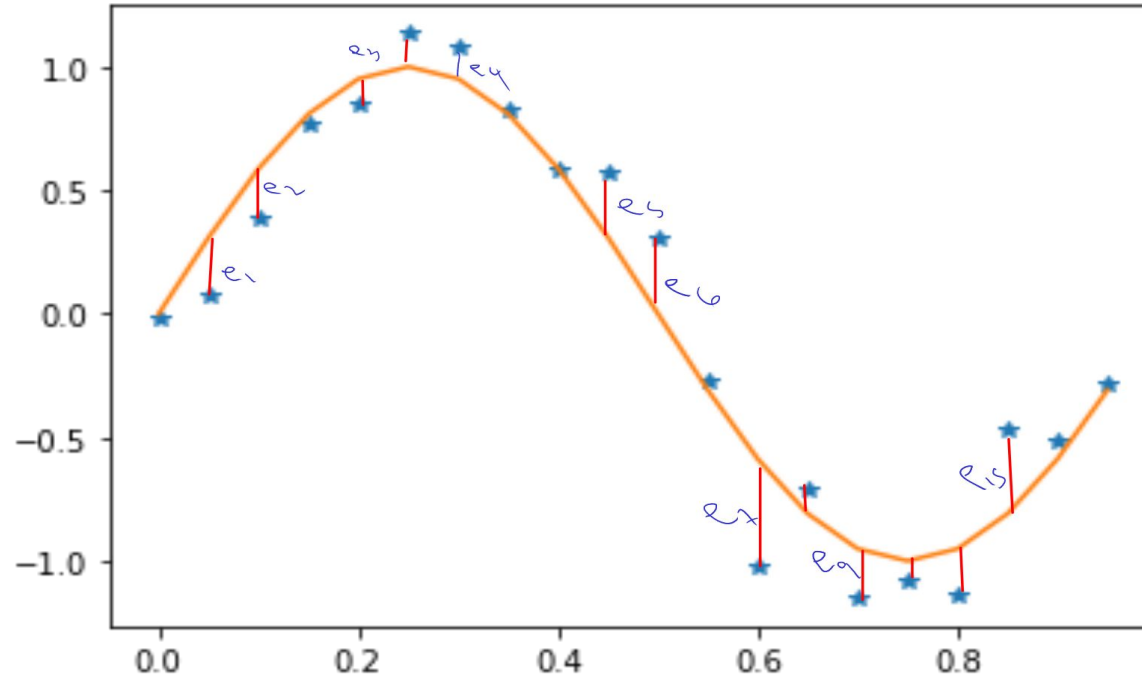
hipótesis:  $h(x_i) = x_i^0 w_0 + \sum_{j=1}^p x_i^j w_j$

hipótesis:  $h(x_i) = \sum_{j=0}^p x_i^j w_j$

$$X \cdot w^T = h(X) = \bar{y}$$

$$h(X) = Xw^t$$

## Función de Pérdida



## Función de Pérdida

$$\mathcal{L} = \frac{\sum_{i=0}^n (y_i - h(x_i))^2}{2n}$$

$$\begin{array}{c}
 y \\
 \hline
 y_1 \\
 y_2 \\
 \vdots \\
 y_i \\
 \vdots \\
 y_n
 \end{array}
 -
 \begin{array}{c}
 \bar{y} \\
 \hline
 \boxed{h(x_1)} \\
 h(x_2) \\
 \vdots \\
 h(x_i) \\
 \vdots \\
 h(x_n)
 \end{array}
 =
 \begin{array}{c}
 y - \bar{y} \\
 \hline
 y_1 - h(x_1) \\
 y_2 - h(x_2) \\
 \vdots \\
 y_i - h(x_i) \\
 \vdots \\
 y_n - h(x_n)
 \end{array}
 =
 \begin{array}{c}
 \vec{e} \\
 \hline
 e_1 \\
 e_2 \\
 \vdots \\
 e_i \\
 \vdots \\
 e_n
 \end{array}$$

$$[e_1 e_2 \dots e_i \dots e_n] \cdot \begin{array}{c} e_1 \\ e_2 \\ \vdots \\ e_i \\ \vdots \\ e_n \end{array} = \sum_{i=0}^n e_i^2 = e^T \cdot e$$

$$\mathcal{L} = \frac{e^t \cdot e}{2n}$$

$$\mathcal{L} = \frac{||Y - h(X)||_2^2}{2n}$$

Normal L2

```
1 def train(x, y, umbral, alfa, p):  
2     np.random.seed(2001)  
3     w = [np.random.rand() for i in range(1,p+1)]  
4     y_pred = h(x,w)  
5     L = Error(y, y_pred)  
6     while (L > umbral):  
7         dw = derivada(x, y, w)  
8         w = update(w,db, alfa)  
9         y = h(x,w)  
10        L = Error(y, y_pred)  
11    return w,L
```



```
1 def train(x, y, umbral, alfa, p):  
2     np.random.seed(2001)  
3     w = [np.random.rand() for i in range(1,p+1)]  
4     y_pred = h(x,w)  
5     L = Error(y, y_pred)  
6     while (L > umbral):  
7         dw = derivada(x, y, w)  
8         w = update(w,db, alfa)  
9         y = h(x,w)  
10        L = Error(y, y_pred)  
11    return w,L
```

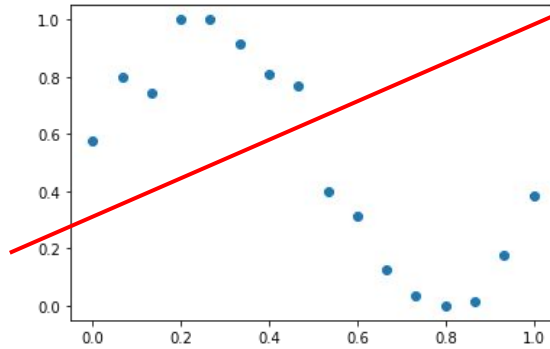
```
1 def train(x, y, umbral, alfa, p):  
2     np.random.seed(2001)  
3     w = [np.random.rand() for i in range(1,p+1)]  
4     y_pred = h(x,w)  
5     L = Error(y, y_pred)  
6     while (L > umbral):  
7         dw = derivada(x, y, w)  
8         w = update(w,db, alfa)  
9         y = h(x,w)  
10        L = Error(y, y_pred)  
11    return w,L
```

```
1 def train(x, y, umbral, alfa, p):  
2     np.random.seed(2001)  
3     w = [np.random.rand() for i in range(1,p+1)]  
4     y_pred = h(x,w)  
5     L = Error(y, y_pred)  
6     while (L > umbral):  
7         dw = derivada(x, y, w)  
8         w = update(w,db, alfa)  
9         y = h(x,w)  
10        L = Error(y, y_pred)  
11    return w,L
```

```
1 def train(x, y, umbral, alfa, p):  
2     np.random.seed(2001)  
3     w = [np.random.rand() for i in range(1,p+1)]  
4     y_pred = h(x,w)  
5     L = Error(y, y_pred)  
6     while (L > umbral):  
7         dw = derivada(x, y, w)  
8         w = update(w,dw, alfa)  
9         y = h(x,w)  
10        L = Error(y, y_pred)  
11    return w,L
```

# 2 Regularización

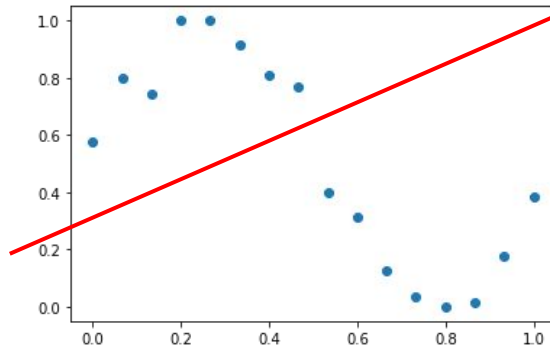
## Underfitting



$$h(x_i) = x_i w + b$$

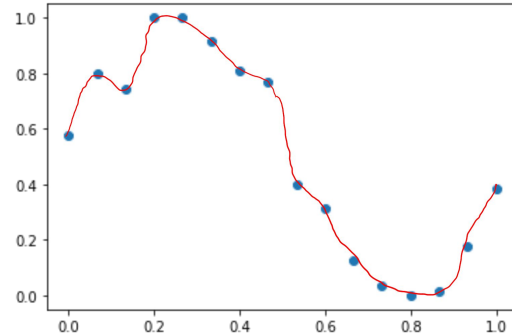


## Underfitting



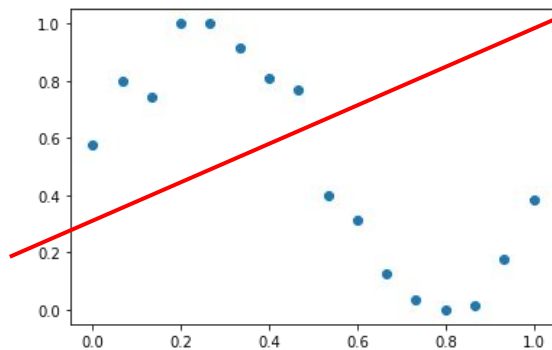
$$h(x_i) = x_i w + b$$

## Overfitting



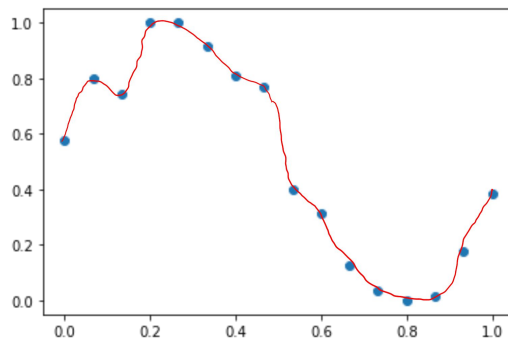
$$h(x_i) = x_i^0 w_0 + x_i^1 w_1 + \dots + x_i^{20} w_{20}$$

Underfitting



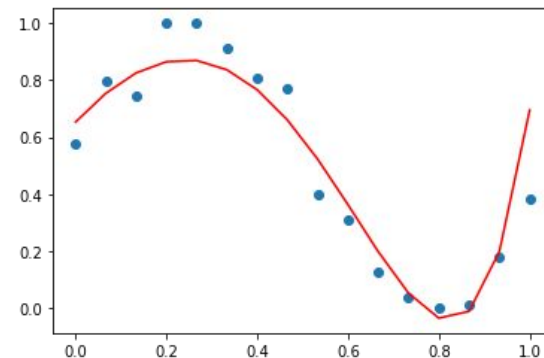
$$h(x_i) = x_i w + b$$

Overfitting



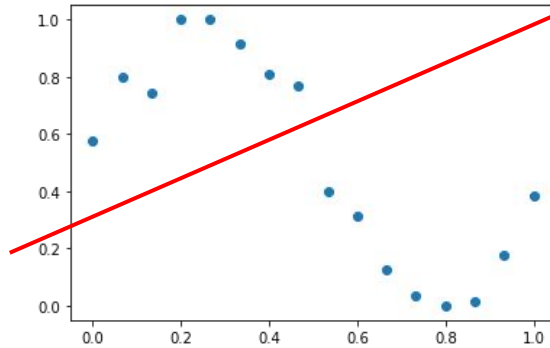
$$h(x_i) = x_i^0 w_0 + x_i^1 w_1 + \dots + x_i^{20} w_{20}$$

good



$$h(x_i) = x_i^0 w_0 + x_i^1 w_1 + \dots + x_i^3 w_3$$

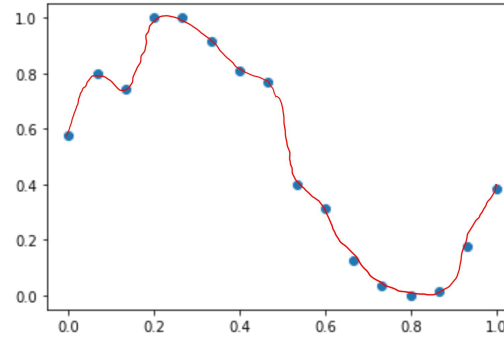
Underfitting



$$h(x_i) = x_i w + b$$

- Modelo simple
- Modelo con poca capacidad

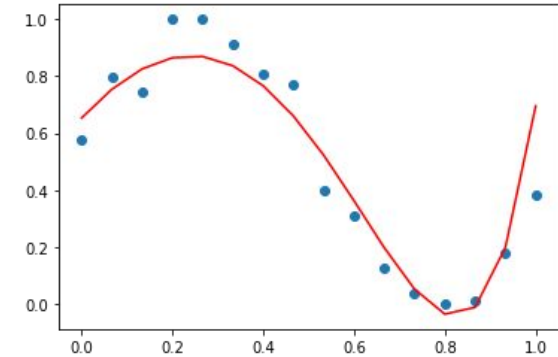
Overfitting



$$h(x_i) = x_i^0 w_0 + x_i^1 w_1 + \dots + x_i^{20} w_{20}$$

- Modelo muy complejo
- Modelo con mucha capacidad

good



$$h(x_i) = x_i^0 w_0 + x_i^1 w_1 + \dots + x_i^3 w_3$$

- Modelo que se ajusta a los datos
- Modelo con capacidad adecuada

¿Cómo hacemos para disminuir la complejidad de nuestro modelo?

# Regularización

Término regularizador

$$\mathcal{L} = \frac{||Y - h(X)||_2^2}{2n} + \mathcal{R}(w)$$

$$\mathcal{L} = \frac{\sum_{i=0}^n (y_i - h(x_i))^2}{2n} + \frac{\lambda}{n} \sum_{j=1}^p w_j^2$$

$$\mathcal{L} = \min \left\{ \frac{\sum_{i=0}^n (y_i - h(x_i))^2}{2n} + \frac{\lambda}{n} \sum_{j=1}^p w_j^2 \right\}$$



$$\mathcal{L} = \frac{\sum_{i=0}^n (y_i - h(x_i))^2}{2n} + \lambda \sum_{j=1}^p w_j^2$$

$$\mathcal{L} = \min \left\{ \frac{\sum_{i=0}^n (y_i - h(x_i))^2}{2n} + \lambda \sum_{j=1}^p w_j^2 \right\}$$

$$\mathcal{L} = \frac{\|Y - XW^t\|_2^2}{2n} + \lambda \|W\|_2^2$$

¿Qué ocurre si el parámetro  $\lambda$  es muy grande ?

$$\lambda = 10000$$

$$\mathcal{L} = \frac{\|Y - XW^t\|_2^2}{2n} + \lambda \|W\|_2^2$$

¿Qué ocurre si el parámetro  $\lambda$  es muy pequeño ?

$$\lambda = 0.01$$

$$\mathcal{L} = \frac{\|Y - XW^t\|_2^2}{2n} + \lambda \|W\|_2^2$$

¿A quién afecta o quienes cambian si modificamos la función de pérdida?

# Las Derivadas

$$\frac{\partial \mathcal{L}}{\partial w_i} = \frac{\partial \left( \frac{\|Y - XW^t\|_2^2}{2n} + \lambda \|W\|_2^2 \right)}{\partial w_i}$$

Hipótesis  $h(X) = X * w^T$

Loss  $\mathcal{L} = \frac{\|Y - XW^t\|_2^2}{2n} + \frac{\lambda}{n} \|W\|_2^2$

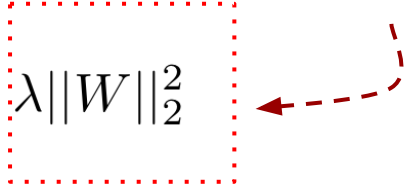
Derivadas  $\frac{\partial \mathcal{L}}{\partial w_j} = \frac{\sum_{i=0}^n (y_i - h(x_i)) * (-x_i^j)}{n} + \frac{2\lambda w_j}{n}$

## 2.1 Métodos de Regularización en Regresión

Loss

$$\mathcal{L} = \frac{\|Y - XW^t\|_2^2}{2n} + \lambda \|W\|_2^2$$

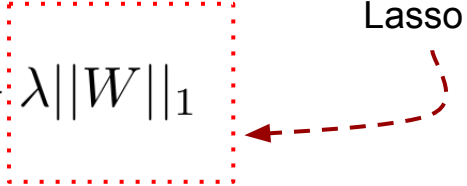
Ridge

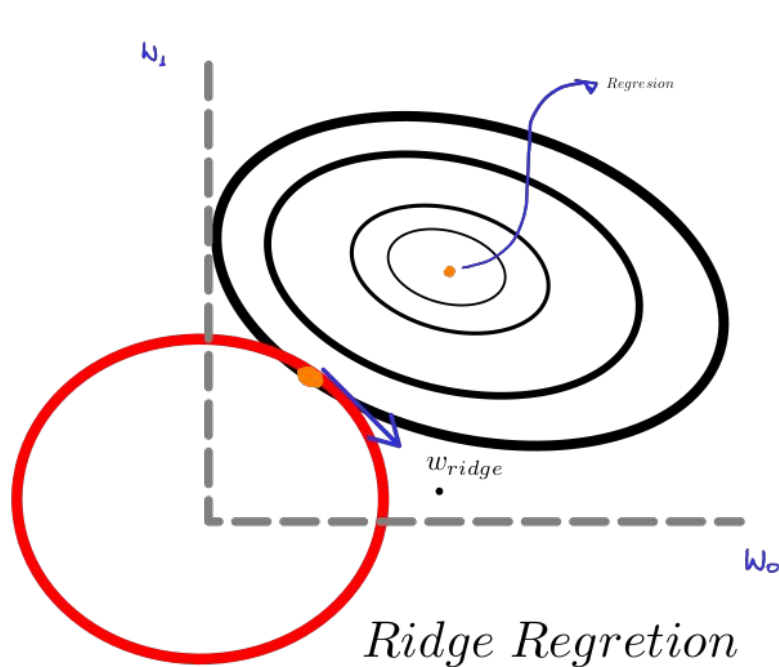


Loss

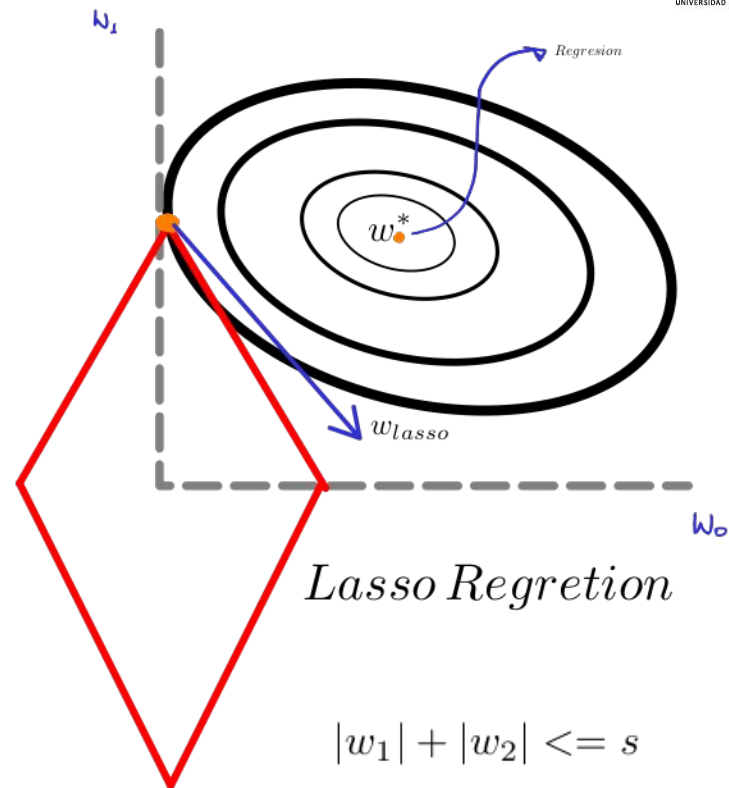
$$\mathcal{L} = \frac{\|Y - XW^t\|_2^2}{2n} + \lambda \|W\|_1$$

Lasso





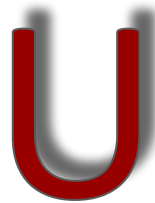
$$w_1^2 + w_2^2 \leq s$$



$$|w_1| + |w_2| \leq s$$

$$\mathcal{L} = \frac{\|Y - XW^t\|_2^2}{2n} + \lambda \|W\|_2^2$$

Valores pequeños de  $W$



$$\mathcal{L} = \frac{\|Y - XW^t\|_2^2}{2n} + \lambda \|W\|_1$$

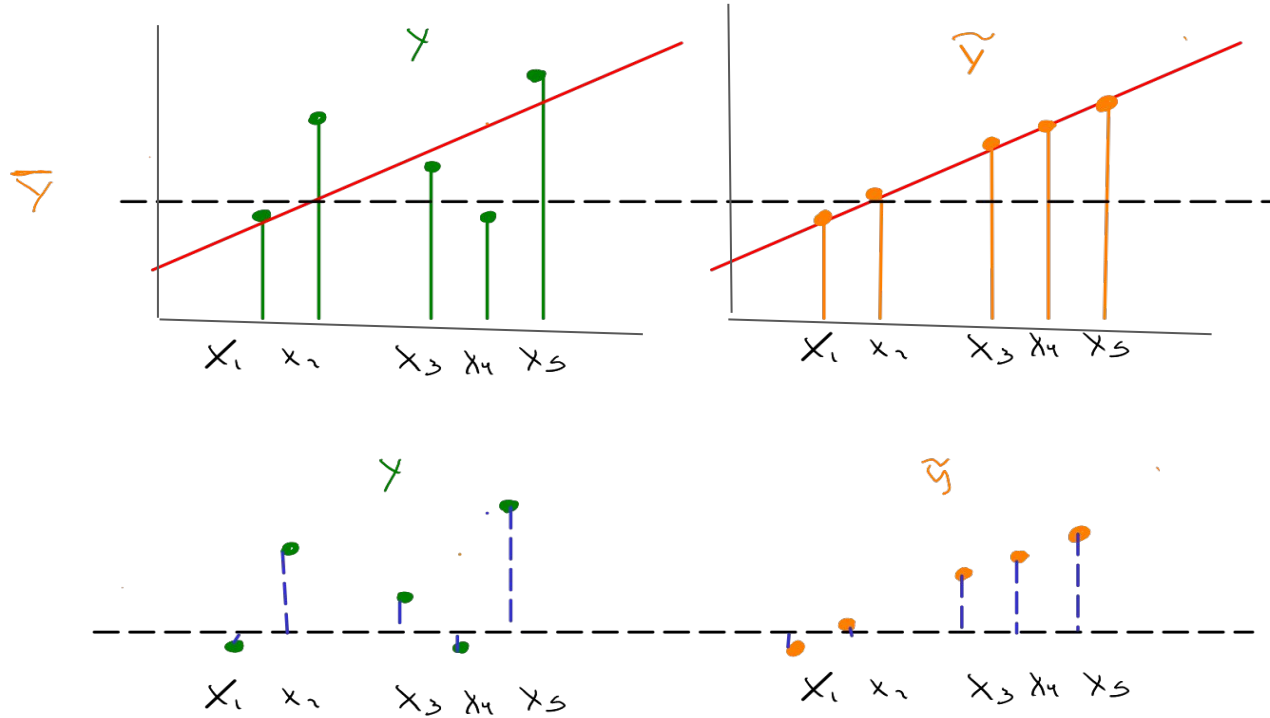
Vector con varios valores cero

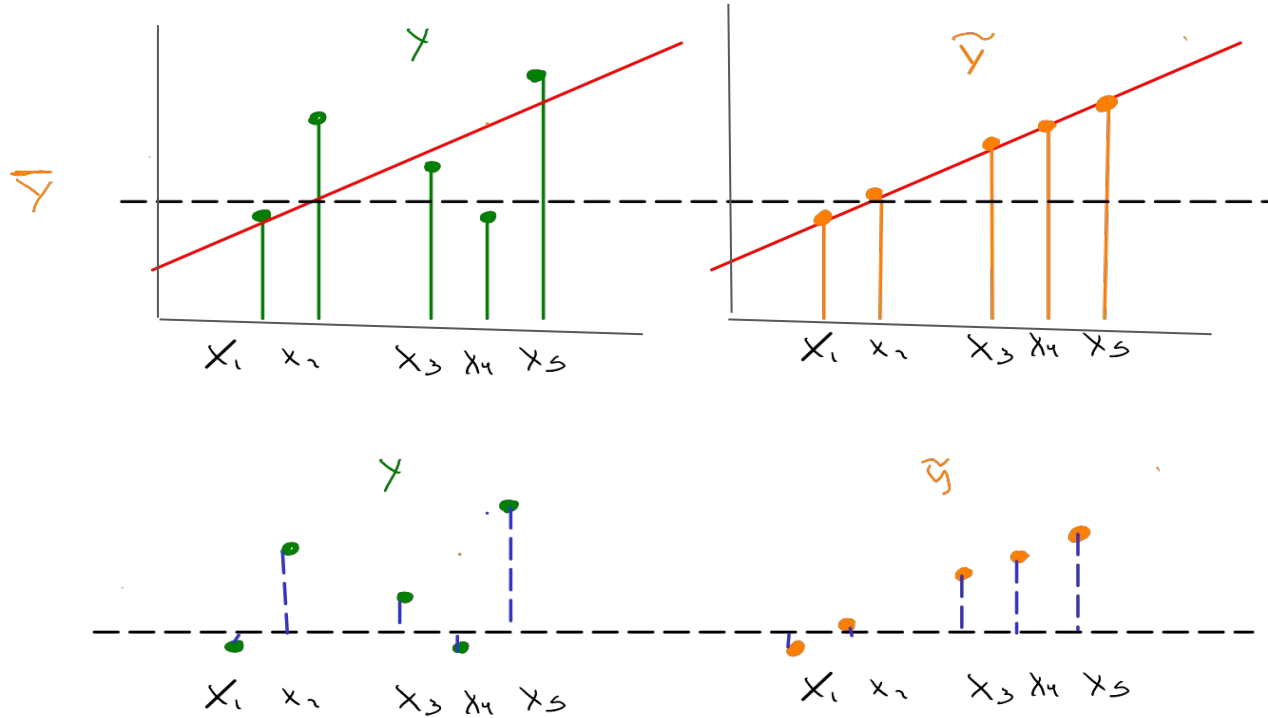


# Elastinet

$$\mathcal{L} = \frac{||Y - XW^t||_2^2}{2n} + \rho\lambda||W||_1 + (1 - \rho)\lambda||W||_2$$

## 2.3 Medida de calidad de la regresión





Coefficiente de  
determinación

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{Y})^2}{\sum_{i=1}^n (y_i - \bar{Y})^2}$$

# 3 Entrenamiento, bias y varianza

Data Set

Training data

Testing data

80%	20%
-----	-----

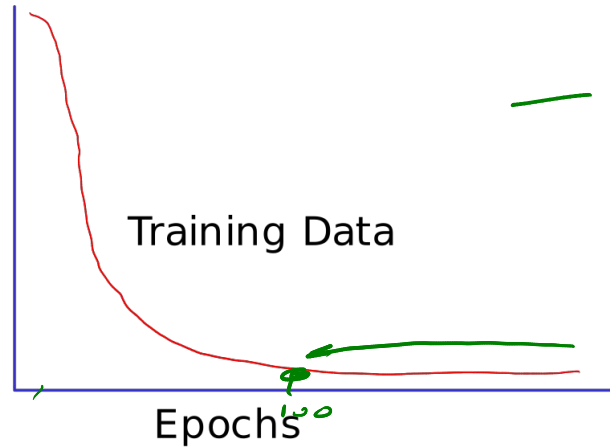
loss, w = training(x\_train, y\_training, alpha, epochs)

while ( $\epsilon > w_{err}$ )

for (i = 1; i < epochs)

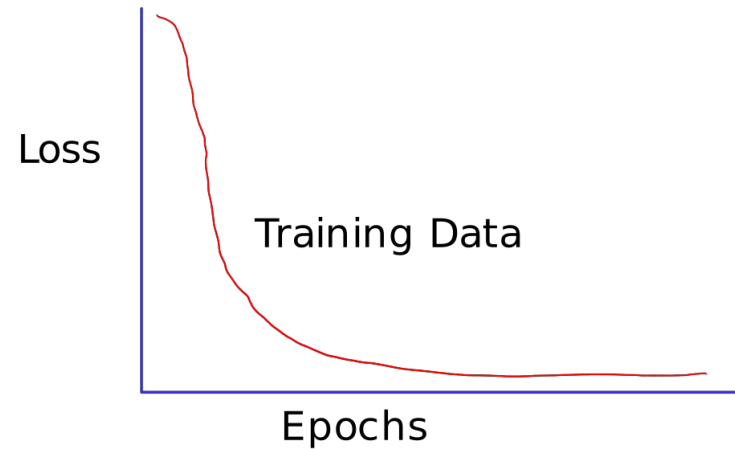
Loss

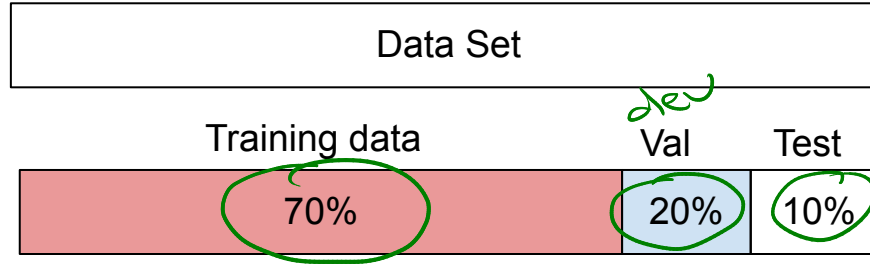
Training Data



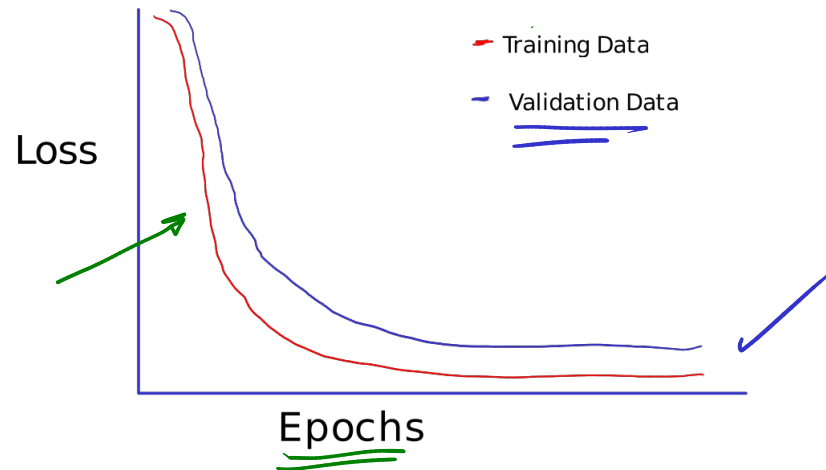
Data Set		
Training data	Val	Test
70%	20%	10%

`loss, w = training(x_train, y_train, alpha, epochs)`





`loss, w = training(x_train, y_train, alpha, epochs)`





```
training(x_train, y_train, alpha, epochs):
```

```
    w = # Initialize parameters
```

```
    for i in range(epochs):
```

```
        x,y = get_data(x_train, y_train, batch)
```

```
        dw = derivatives(x,y,w)
```

```
        w = Change_Parameters(w,dw, alpha)
```

```
        loss_training = Error(x,y,w,batch)
```

```
        loss_val = Error(x_val,y_val,w,batch)
```

```
        L_T.append(loss_training)
```

```
        L_V.append(loss_val)
```

```
    return L_T, L_V, w
```

Todos

• SD

• SGD

• MB

10<sup>9</sup>

• 21  
• 16

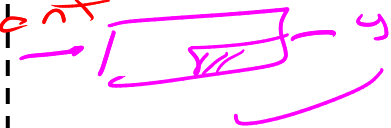


gradient descendant

stochastic gradient descendant

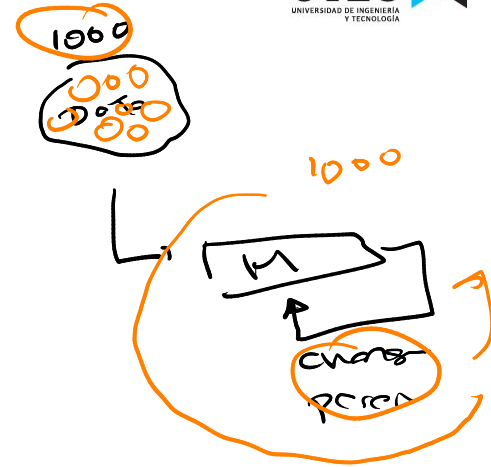
mini batch

$$= \sum_{i=0}^{10^6} \frac{(y_i - \bar{y}_i)^2}{2}$$



```

training(x_train, y_train, alpha, epochs):
    w = # Initialize parameters
    for i in range(epochs):
        x,y = get_data(x_train, y_train, batch)
        dw = derivatives(x,y,w)
        w = Change_Parameters(w,dw, alpha)
        loss_training = Error(x,y,w,batch)
        loss_val = Error(x_val,y_val,w,batch)
        L_T.append(loss_training)
        L_V.append(loss_val)
    return L_T, L_V, w
    
```



```

training(x_train, y_train, alpha, epochs):

```

```

    w = # Initialize parameters

```

```

    for i in range(epochs):

```

```

        x,y = get_data(x_train, y_train, batch)

```

```

        dw = derivatives(x,y,w)

```

```

        w = Change_Parameters(w,dw, alpha)

```

```

        loss_training = Error(x,y,w,batch)

```

```

        loss_val = Error(x_val,y_val,w,batch)

```

```

        L_T.append(loss_training)

```

```

        L_V.append(loss_val)

```

```

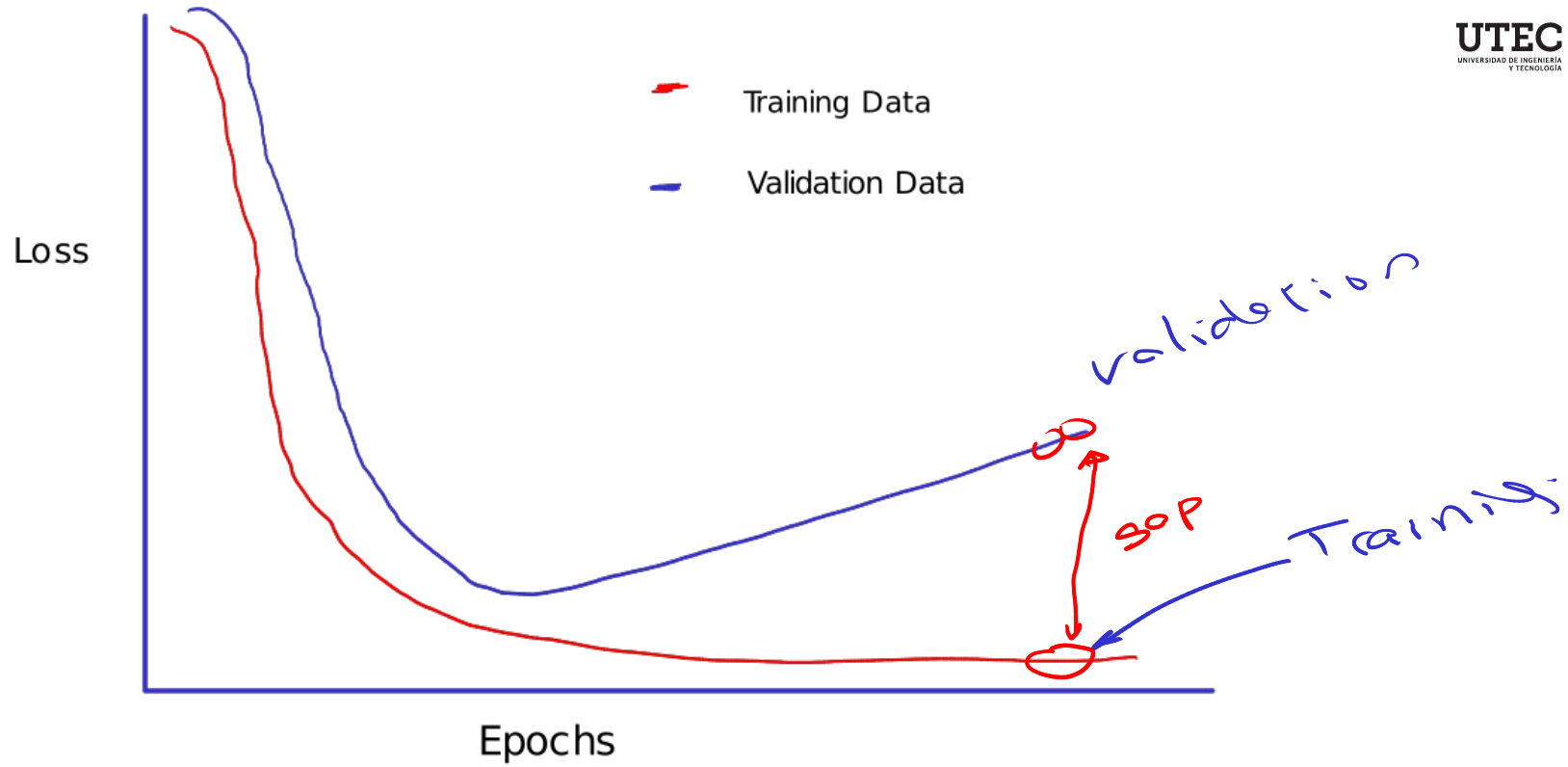
    return L_T, L_V, w

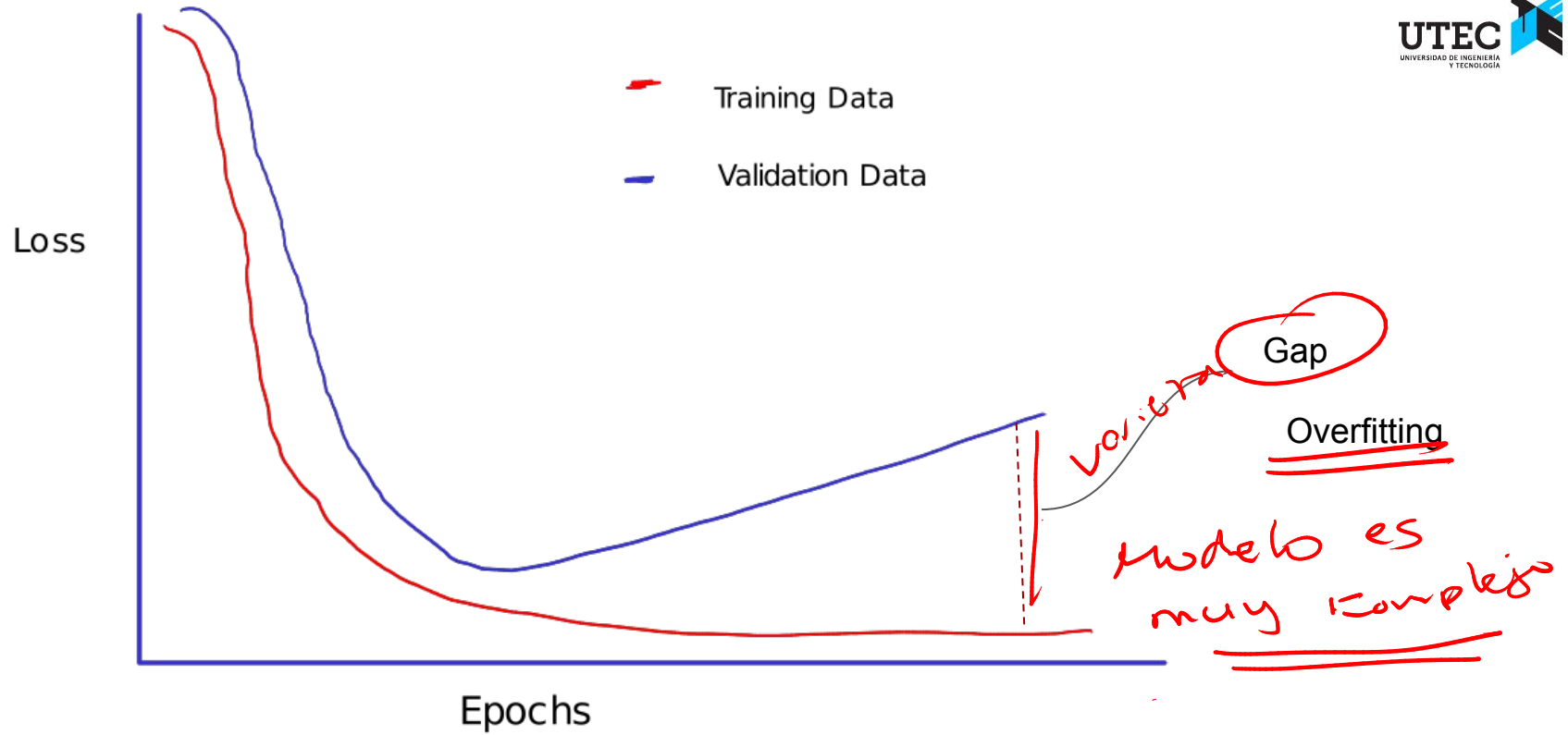
```

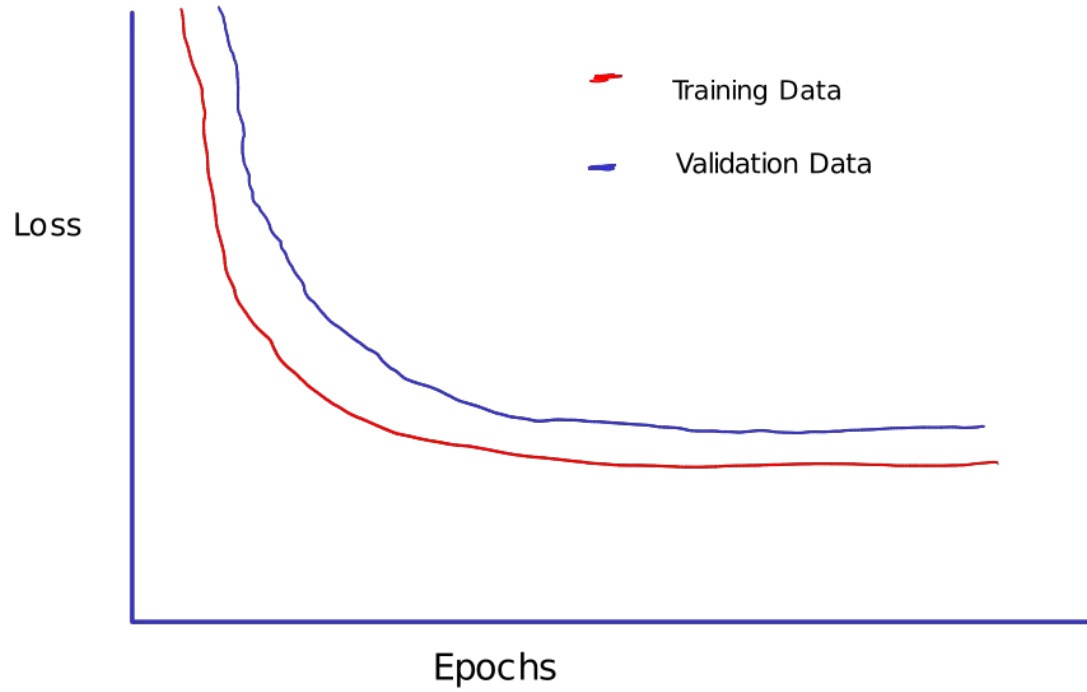
```
traing(x_train, y_train, alpha, epochs):  
    w = # Initialize parameters  
    for i in range(epochs):  
        x,y = get_data(x_train, y_train, batch)  
        dw = derivatives(x,y,w)  
        w = Change_Parameters(w,dw, alpha)  
        loss_training = Error(x,y,w,batch)  
        loss_val = Error(x_val,y_val,w,batch)  
        L_T.append(loss_training)  
        L_V.append(loss_val)  
    return L_T, L_V, w  
  
plot(loss_t, loss_v, epochs)
```

```
training(x_train, y_train, alpha, epochs):  
    w = # Initialize parameters  
    for i in range(epochs):  
        x,y = get_data(x_train, y_train, batch)  
        dw = derivatives(x,y,w)  
        w = Change_Parameters(w,dw, alpha)  
        loss_training = Error(x,y,w,batch)  
        loss_val = Error(x_val,y_val,w,batch)  
        L_T.append(loss_training)  
        L_V.append(loss_val)  
    return L_T, L_V, w
```

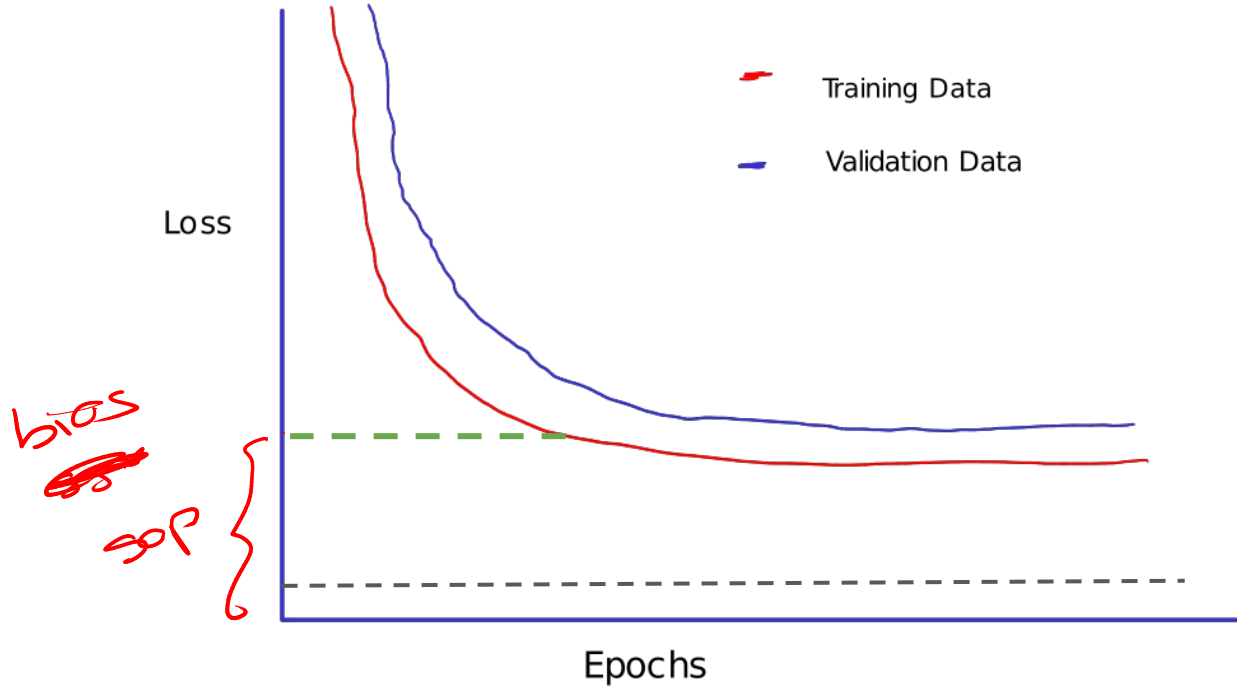
```
plot(loss_t, loss_v, epochs)
```

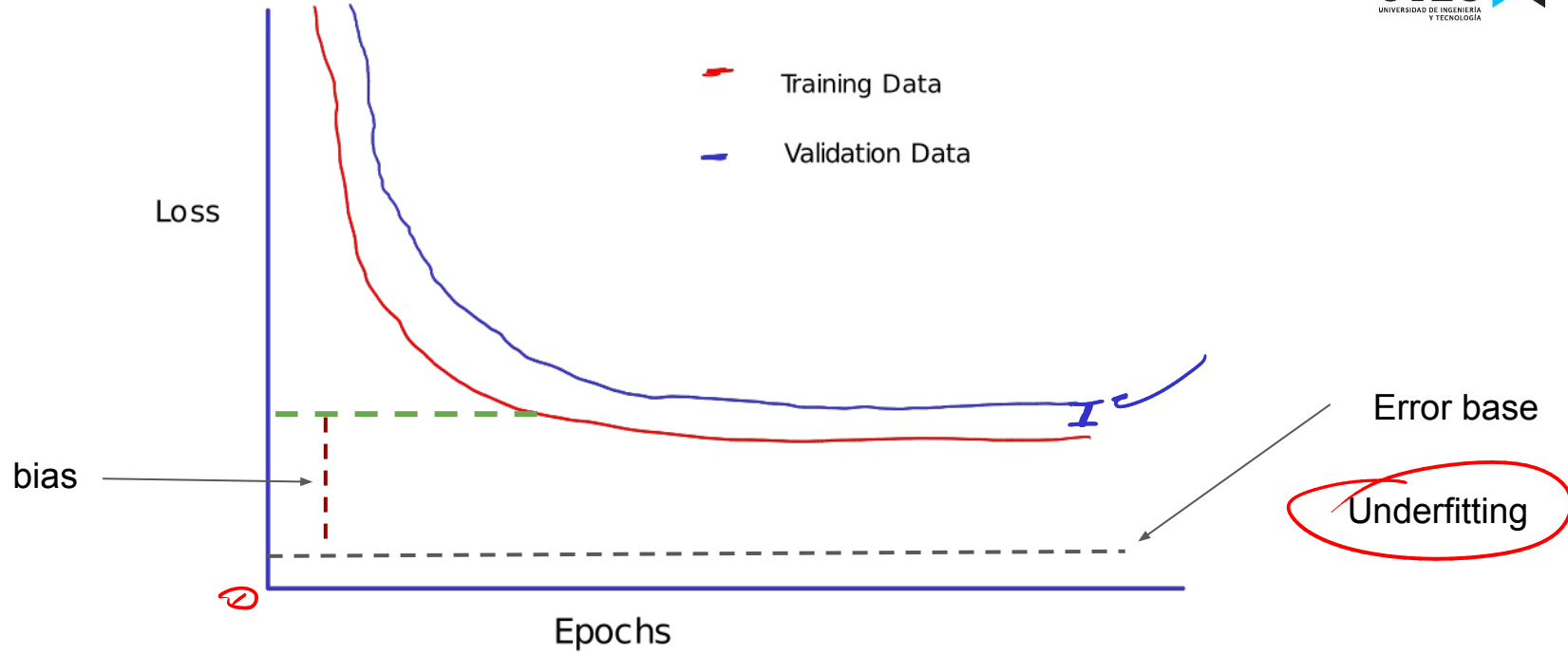


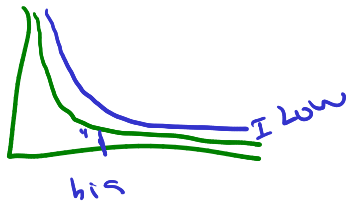










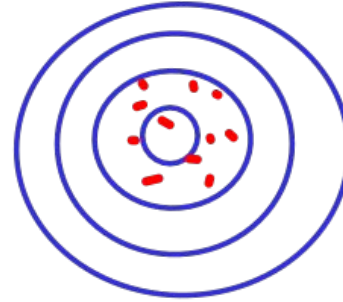
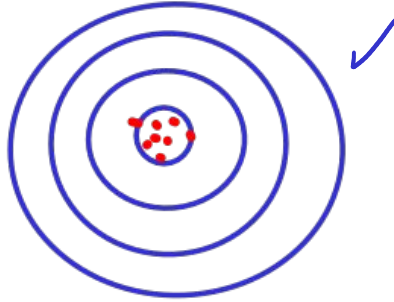


Low Variance a)

High Variance b)

Low bias

El modelo es  
consistente y  
preciso

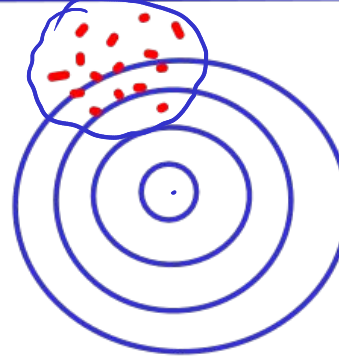
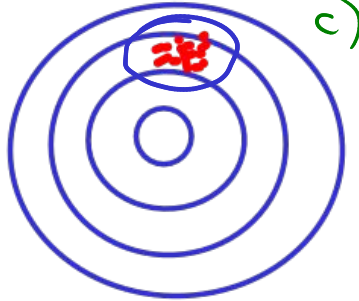


El modelo es incierto  
pero preciso.

overfitting

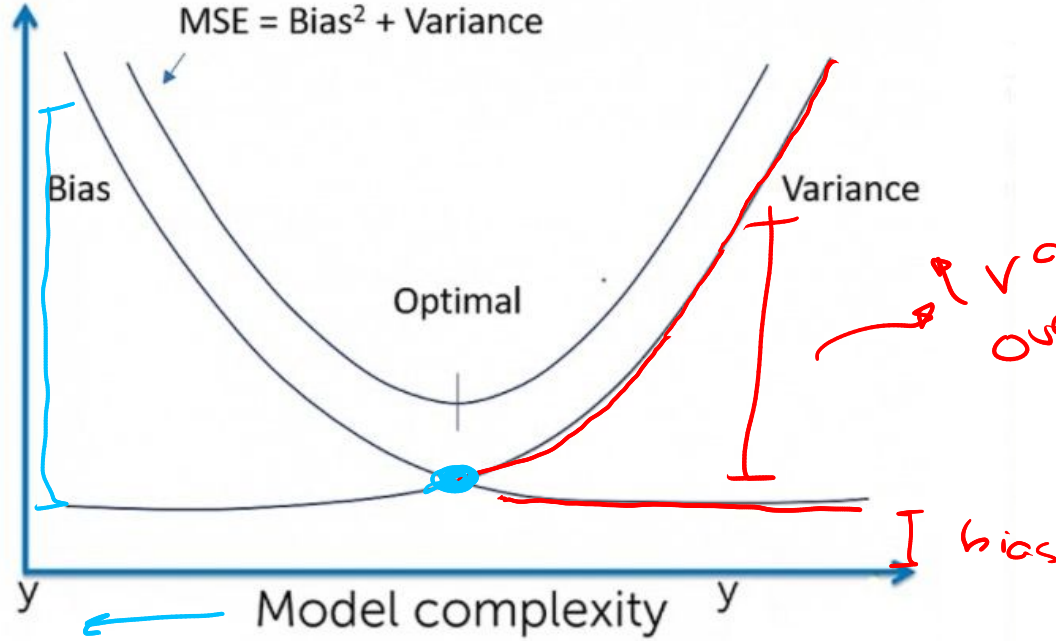
High Bias

El modelo es  
consistente  
pero inexacto.



El modelo es incierto  
e inexacto en  
promedio

Loss



Fuente: <https://editor.analyticsvidhya.com/uploads/983161.png>

# Clasificación

Cristian López Del Alamo

[clopezd@utec.edu.pe](mailto:clopezd@utec.edu.pe)

IPRODAM3D - Research group

2022

Clasificación

Supervisi



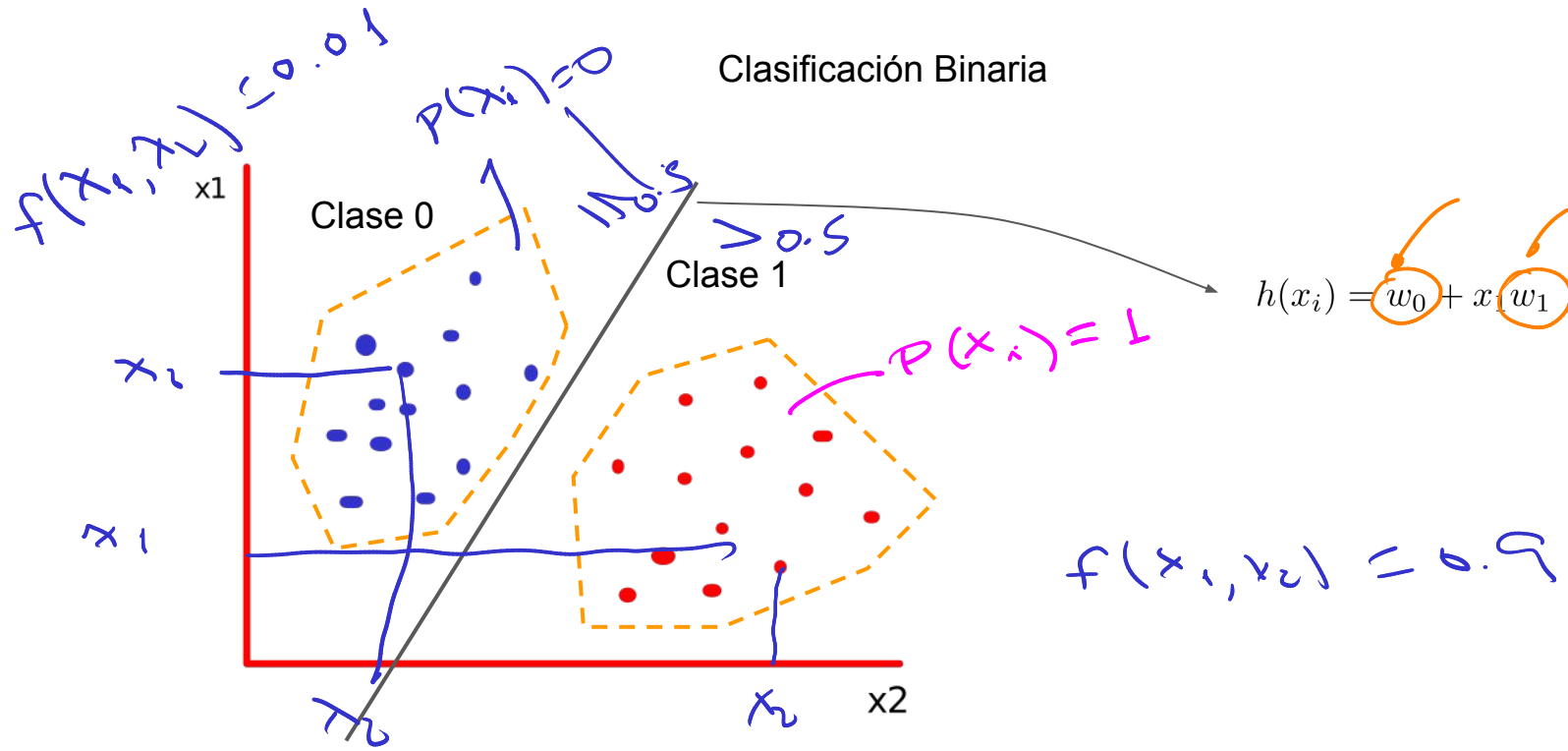
$(x, y)$   
 $x$   
 $y - y$



forming 1 2 To merge

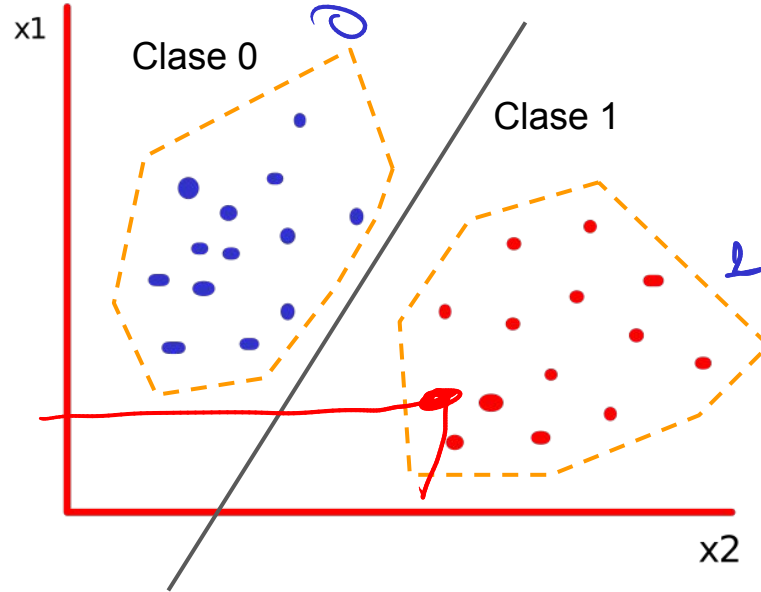
$x_1$		
$x_2$		
.		
.		
.		
.		
.		
$x_n$		

# Clasificación Binaria





# Clasificación Binaria



$$h(x_i) = w_0 + x w_1$$

$$s(x_i) = \frac{1}{1 + e^{-h(x_i)}}$$

$$s(x_1, x_2) = 0.2$$
$$s(x_1, x_2) = 0.85$$

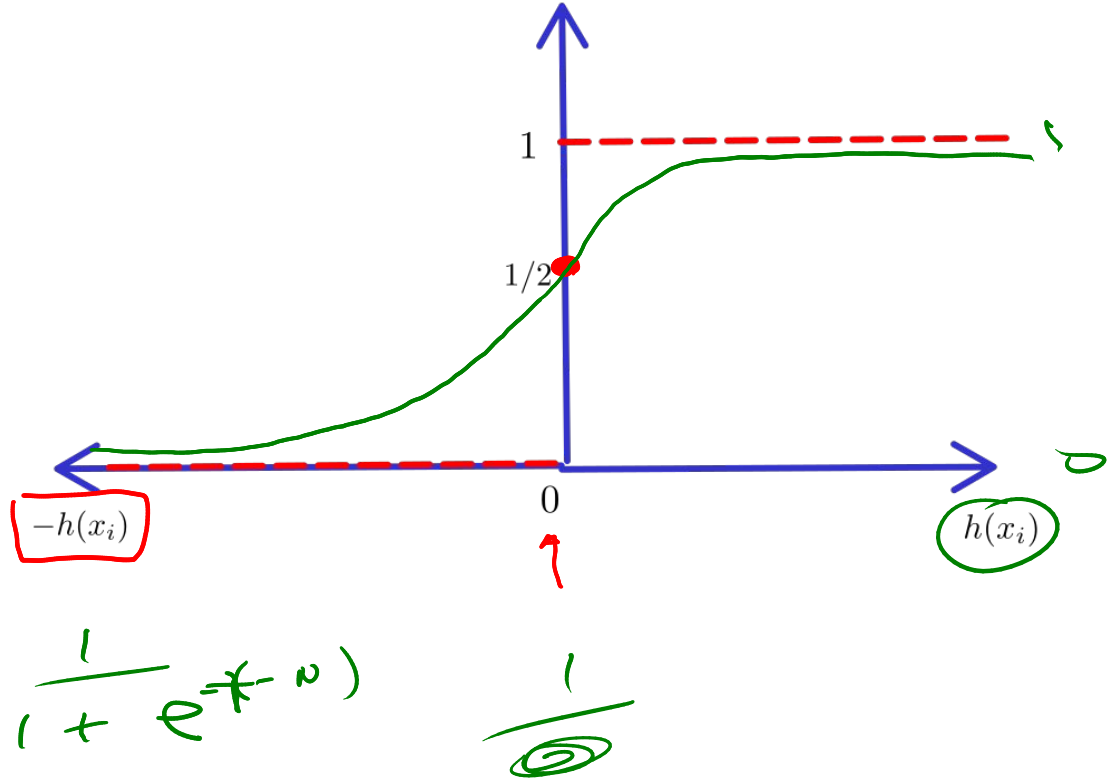
# Clasificación Binaria

$$x_i = \langle x_{i1}, x_{i2} \rangle$$

$$s(x_i)$$

$$s(x_i) = \frac{1}{1 + e^{-h(x_i)}}$$

$$\frac{1}{1 + e^{-h(x_i)}}$$



Hipótesis  $s(x_i) = \frac{1}{1 + e^{-h(x_i)}}$

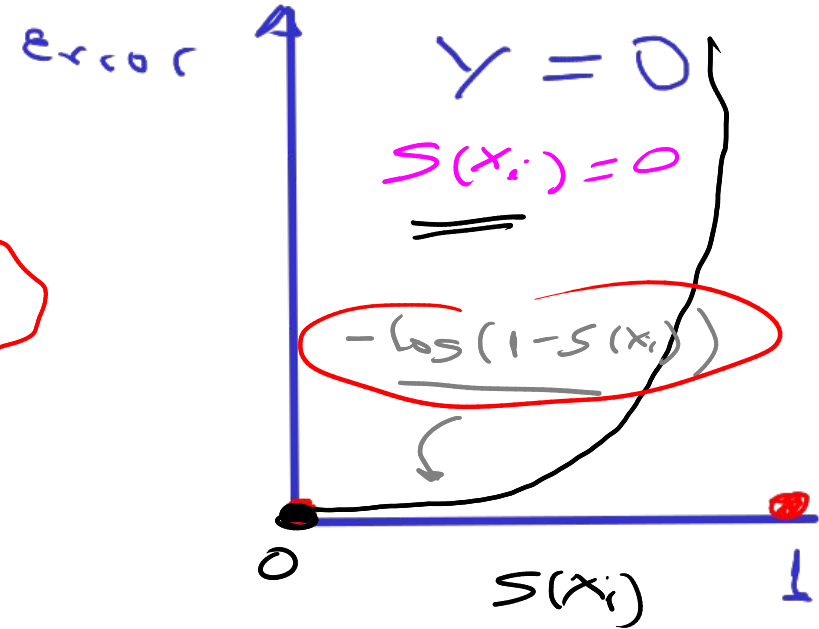
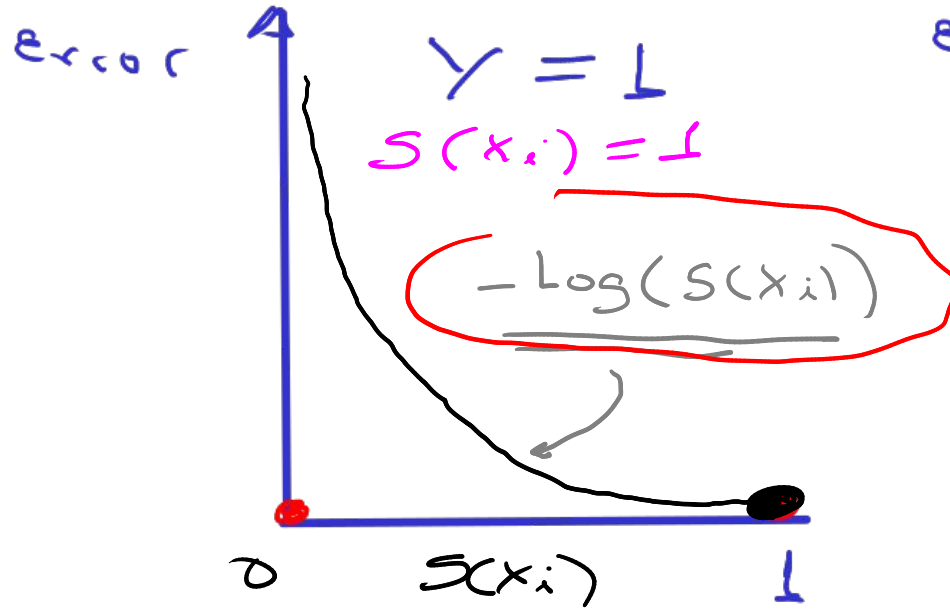
$s(x_i) = [0, 1]$

$y_i = 1 \quad s(x_i) = 0.1$

Loss  $\mathcal{L} = - \sum_{i=1}^n \left[ \underbrace{y_i \log(s(x_i))}_{(1)} + \underbrace{(1 - y_i) \log(1 - s(x_i))}_{(2)} \right]$

(1)  $y = 1$

(2)  $y = 0$

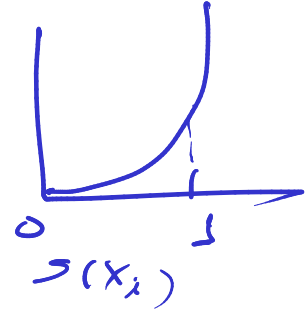


Hipótesis  $s(x_i) = \frac{1}{1 + e^{-h(x_i)}}$

$y = 1$

$S(z_i) = 0.7$

$y = 0$   
 $S(x_i) = 0.9$



Loss  $\mathcal{L} = - \sum_{i=1}^n (y_i \log(s(x_i)) + (1 - y_i) \log(1 - s(x_i)))$

$(-\log(s(x_i)))$

$(-\log(1 - s(x_i)))$

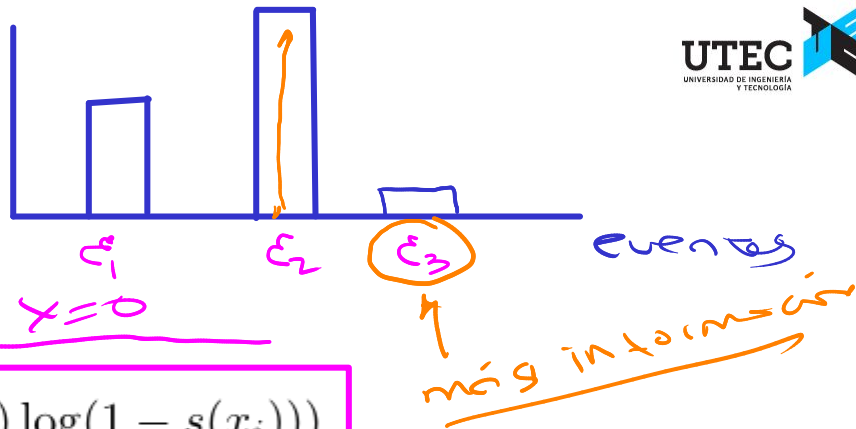
Derivadas  $\frac{\partial L}{\partial w_j} = \frac{1}{n} \sum_{i=1}^n (y_i - s(x_i))(-x_{ij})$

$-\log(p(x_i))$



$e = 0.0002$   $s(x_i) + e$   $\text{frac}$

Hipótesis  $s(x_i) = \frac{1}{1 + e^{-h(x_i)}}$



Loss  $\mathcal{L} = - \sum_{i=1}^n (y_i \log(s(x_i)) + (1 - y_i) \log(1 - s(x_i)))$

Derivadas  $\frac{\partial L}{\partial w_j} = \frac{1}{n} \sum_{i=1}^n (y_i - s(x_i)) (-x_{ij})$

$\log \left( \frac{1}{P(x)} \right)$   $\log 1$

$- \log(P(x))$

$\log_2 \left( \frac{1}{2^{-50}} \right)$

Hipótesis  $s(x_i) = \frac{1}{1 + e^{-h(x_i)}}$

Loss  $\mathcal{L} = - \sum_{i=1}^n (y_i \log(s(x_i)) + (1 - y_i) \log(1 - s(x_i)))$

Derivadas  $\frac{\partial L}{\partial w_j} = \frac{1}{n} \sum_{i=1}^n (y_i - s(x_i))(-x_{ij})$

Hipótesis  $s(x_i) = \frac{1}{1 + e^{-h(x_i)}}$

Loss  $\mathcal{L} = - \sum_{i=1}^n (y_i \log(s(x_i)) + (1 - y_i) \log(1 - s(x_i)))$





INGENIERIA  
MECATRONICA

BIOTECNOLOGIA

CIENCIA DE  
LA COMPUTACION

INGENIERIA  
AMBIENTAL

INGENIERIA  
ENERGIA

INDUSTRIAL

ELECTRONICA



UTEC

UNIVERSIDAD DE INGENIERIA  
Y TECNOLOGIA

